



Analyse linguistique de documents d'entreprises : demande appliquée et théorisation des phénomènes

Anne Condamines

► To cite this version:

Anne Condamines. Analyse linguistique de documents d'entreprises : demande appliquée et théorisation des phénomènes. D. Alamargot, P. Terrier et J.-M. Cellier. Production, compréhension et usages des écrits techniques au travail, Octares, pp.17-30, 2005. <halshs-00814841>

HAL Id: halshs-00814841

<https://halshs.archives-ouvertes.fr/halshs-00814841>

Submitted on 17 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyse linguistique de documents d'entreprises : demande appliquée et théorisation des phénomènes

Anne Condamines
Equipe de Recherche en Syntaxe et Sémantique, UMR 5610
CNRS et Université Toulouse Le Mirail
anne.condamines@univ-tlse2.fr

Résumé : Dans une perspective qui refuse d'opposer linguistique appliquée et linguistique théorique, l'article montre comment, à travers l'analyse de corpus, ces deux approches peuvent au contraire se compléter et s'enrichir. La proposition de l'auteur s'inscrit dans une sémantique doublement située : par la situation de production des textes des corpus et par la situation d'interprétation, c'est-à-dire l'objectif de l'étude. Point de vue appliqué et point de vue théorique constituent deux types d'objectifs possibles, également situés dans un état de connaissance, à un moment donné et qui varient essentiellement par les modes d'évaluation et d'intégration de leurs résultats. Plusieurs exemples montrent comment des problématiques appliquées ont donné lieu à une avancée théorique sur la connaissance en linguistique.

Abstract : Within a framework refusing to contrast applied linguistics and theoretic linguistics, this paper shows how these two approaches may complement and enrich one another through corpus analysis. The idea put forward by the author is anchored in a bi-situational semantics : on the one hand, the text production situation and, on the other hand, the interpretation situation, namely the aim of the study. Both applied and theoretic points of view constitute feasible objectives, equally located at a given time in a knowledge state, which mostly differ in the way results are validated and taken into account. The author uses several examples to show how applied problems have given rise to a theoretic advance in the field of linguistics.

1. Introduction

Lorsque le matériau sur lequel travaille le linguiste est constitué d'écrits provenant d'une entreprise, des questions nouvelles se posent sur la méthode à mettre en place, la nature technique de ce matériau ne pouvant être tenue pour caractéristique secondaire. En effet, il est assez rare que ces documents techniques soient arrivés jusqu'au linguiste sans être

accompagnés d'une demande précise à leur sujet. Les entreprises sont suffisamment jalouses de leur documentation pour accepter de la céder sans contre partie, sans que cette externalisation ne soit accompagnée d'une plus-value en termes de connaissances. Pour le linguiste, cette situation qui consiste à répondre à une demande en lien avec l'expression d'un besoin ne va pas de soit. L'objectif de cet article est à la fois d'expliquer ce qui fait la nouveauté de cette situation pour la linguistique, voire son étrangeté, mais aussi de montrer que la nécessité de répondre à une demande peut ouvrir des réflexions et des questionnements et, en fin de compte, enrichir la réflexion théorique en linguistique.

2. « L'application » en linguistique

Dans l'histoire de la linguistique, la prise en compte d'une quelconque application a été longtemps considérée comme ne faisant pas partie de la problématique théorique que souhaitait se construire cette discipline. La pression de la demande sociétale accompagnée du développement d'outils de traitement de la langue a conduit à un changement très net qui demande une véritable mutation épistémologique, encore en cours, de la linguistique. Ce paragraphe montrera quels sont les enjeux et les possibilités du changement.

2.1. L'application dans l'histoire de la linguistique

Une grande partie du XX^e siècle a été marquée, pour la linguistique, par une volonté de construire un objet parfaitement circonscrit, qu'on pourrait étudier de manière autonome, c'est-à-dire un objet, la langue, isolé de tout contexte réel d'utilisation. Dans cette perspective sous tendue par une vision structuraliste, il s'agissait de comprendre ce qui faisait système, de mettre au jour des régularités structurelles stables. Les usages effectifs ne pouvaient être pris en considération et encore moins les applications pour justifier les études menées. Cette situation a perduré jusqu'à il y a une vingtaine d'années. L'application n'était pas totalement niée mais au mieux considérée comme seconde, comme un objectif qui ne pouvait que prendre en compte les résultats de l'étude théorique, c'est-à-dire les « vrais » résultats de la linguistique considérée comme discipline scientifique :

« Par linguistique appliquée, on désigne l'ensemble des recherches qui utilisent les démarches de la linguistique proprement dite pour certains problèmes de la vie courante et professionnelle... Partie utilitaire et pratique de la linguistique, elle est nécessaire mais ne

peut évidemment constituer la fin unique des recherches en matière de langage ». (Dubois et al., 1973).

Les choses ont un peu changé mais les linguistes de terrain¹ sont encore souvent confrontés à des difficultés :

« Participer à un travail de terrain et appréhender des données de langage socialement situées constitue un projet intellectuel qui pour de nombreux linguistes ne fait pas sens ». (Boutet, 1995, 21).

Si la couverture de la discipline linguistique est en train de changer, y compris au cœur même de ce qui la constitue comme science, c'est en grande partie parce que la question des usages réels et de leur prise en compte dans les analyses est arrivée en force sur le devant des préoccupations. Désormais, il n'est plus de domaine (morphologie, syntaxe, sémantique, discours...) qui ne soit pas interrogé par la confrontation avec la dimension de variation que manifestent ces usages. Dans ce nouveau contexte, le problème de l'application prend aussi une toute autre place.

2.2. Irruption de la linguistique de corpus

La question de l'application en linguistique s'est posée de manière radicalement nouvelle depuis que les corpus constituent un matériau d'étude digne d'intérêt pour les linguistes. Cette évolution a été le fruit d'une mutation socio-scientifique très importante et ce changement s'accompagne nécessairement d'interrogations sur le statut du linguiste et sur son implication par rapport au matériau langagier. Plus précisément, un des points centraux de la réflexion se focalise sur le rôle de l'*interprétation* des phénomènes et sur les moyens de comprendre comment elle se met en place.

2.2.1. Une situation socio-scientifique nouvelle

Même pendant la longue période au cours de laquelle la linguistique s'était donnée un objet décontextualisé comme matériau d'étude, plusieurs courants de la linguistique ont continué à

¹ Après avoir longtemps été réservé aux linguistes travaillant sur le lieu d'utilisation d'une langue particulière, le plus souvent parlée par un très petit nombre d'utilisateurs, ce terme (de linguistes de terrain) est maintenant aussi utilisé pour les linguistes qui travaillent sur les lieux d'utilisation de leur langue maternelle et tiennent compte des contextes d'usages ; cet élargissement de la couverture sémantique de ce terme est le signe d'une évolution dans la discipline.

prendre les corpus comme référence pour leurs travaux ; soit par obligation lorsque le linguiste ne pouvait se fier à sa compétence (domaine historique ou domaines experts), soit par choix méthodologique, avec la conviction que la langue ne pouvait être étudiée indépendamment de son contexte social (sociolinguistique (Labov, 1976)). Mais ces approches étaient marginalisées ; ce n'est que récemment, surtout depuis les années 80-90, que la question des corpus s'est posée d'une manière très nette. Cette irruption de la linguistique de corpus peut s'expliquer de plusieurs façons. D'une part sans doute, les possibilités d'une linguistique essentiellement introspective commençaient à atteindre leurs limites. D'autre part, la demande sociétale en matière de traitement de textes réels est devenue très forte². Elle s'est accompagnée du développement d'outils d'analyse de corpus qui ont eu besoin, très rapidement, d'être alimentés par des ressources obtenues grâce à des analyses de corpus et ont conduit les chercheurs à s'interroger sur le rôle des corpus (Péry-Woodley, 1995). Dans ces années 90 en effet, le TAL (Traitement Automatique de la Langue) a subi une évolution qui l'a mené d'un point de vue où il s'agissait surtout de vérifier la cohérence de théories formelles à un autre point de vue, où il s'agit de traiter réellement des corpus pour des objectifs divers : recherche d'information, extraction d'information, veille scientifique mais aussi construction de connaissances linguistiques à partir de corpus (morphologiques, syntaxiques, sémantiques) pour améliorer les résultats produits par les outils (Pierrel, 2000). Cette évolution ne pouvait qu'interpeller les linguistes qui partagent avec les chercheurs en TAL le même matériau d'étude. De fait, la communauté des linguistes est désormais confrontée au problème majeur que pose la prise en compte de corpus, celui de la variation des phénomènes par opposition à la vision « figée » de l'approche structuraliste.

Une façon de neutraliser le problème de la variation consiste à ne considérer les textes que comme des réservoirs d'attestations : un fait langagier avéré confirme une intuition ou permet de démarrer une réflexion sur ce qui fait système, indépendamment du contexte extralinguistique dans lequel il a été produit. Mais ce point de vue est largement remis en question à la fois par les analystes de discours plus enclins à prendre en compte le contexte de production mais aussi par des linguistes intéressés par des problèmes plus structurels. Il est maintenant acquis par la majorité des chercheurs qu'un corpus n'est pas un tas de textes mais un ensemble de textes rassemblés pour étudier une hypothèse voire répondre à une demande applicative (Habert et al., 1997). Dans cette perspective, la nature des textes est évidemment

² L'accroissement de cette demande est pour beaucoup venu d'un besoin en gestion de la documentation dont le volume (spécifications, manuels, maintenance...) est devenu démesuré. Par exemple, on considère que la documentation concernant un avion contiendrait en volume papier à l'intérieur de l'avion.

prise en compte. Par ailleurs, ce point de vue supprime, en partie du moins, le fossé qui pouvait exister entre visée théorique et visée applicative. En effet, étant donné que le point de focalisation de l'analyse est constitué par le corpus, la nature théorique ou appliquée de l'approche que l'on adopte correspond à deux points de vue certes différents mais pas incompatibles, j'y reviendrai.

2.2.2. Le problème de l'interprétation

Dans le cas d'une linguistique introspective, les exemples sont fournis par l'analyste lui-même comme autant de données avérées qu'il convient d'étudier comme s'il s'agissait de faits « naturels » (Corbin, 1980). Dans une vision idéale de la discipline, l'analyse est alors menée d'une manière « neutre », en tout cas qui se revendique comme telle. Mais de nombreux chercheurs (Foucault, 1966), (Auroux, 1998) ont montré que cette prétendue neutralité était surtout gouvernée par la recherche d'une dimension scientifique qui ferait de la linguistique une discipline différente des autres sciences humaines, plus proche des sciences dites dures. La confrontation avec la réalité des productions langagières oblige à revenir sur cet espoir de neutralité. Dès que l'on travaille sur des textes en effet, on est obligé de faire le constat que l'on construit un sens plus qu'on ne retrouve ce sens (Boutet, 1995): il ne s'agit pas de reconstituer ce qu'a voulu dire l'auteur. La situation d'analyse d'un corpus pour un linguiste présente deux caractéristiques majeures :

- d'une part, le linguiste n'est pas le destinataire des textes qu'il étudie (sauf cas exceptionnel où il a mené des entretiens) et il ne « lit » pas ces textes avec l'objectif pour lequel ils ont été écrits ;
- d'autre part, l'interprétation se fait toujours en différé : le linguiste ne dispose ainsi jamais de toutes les conditions d'énonciation dans lesquelles s'est élaboré le texte étudié.

La question alors est de comprendre quels sont les éléments qui contribuent à la construction de ce sens, c'est-à-dire à l'interprétation des faits langagiers.

2.3. Une sémantique doublement située.

Finalement, si le sens n'est pas un donné du corpus qu'il faut retrouver mais bien un construit, la question est de savoir comment on peut aller vers une systématisation de l'interprétation. Hormis les phénomènes linguistiques du corpus, deux éléments majeurs interviennent dans l'interprétation :

- la situation de production des textes : statut du (des) locuteur(s) et du (des) destinataire(s), objectif de la rédaction, genre du texte...
- la situation d'interprétation : objectif de l'analyse.

L'analyse qui est menée est ainsi « doublement située » et c'est la réflexion sur ce double balisage qui pourrait permettre de donner une assise théorique aux études menées sur corpus. Ainsi présentée, la sémantique textuelle peut relever aussi bien d'une approche appliquée que d'une approche théorique.

2.3.1. De l'application à la théorisation

Il convient à présent de s'interroger sur les rapports entre objectif appliqué et objectif théorique, étant bien entendu que l'un comme l'autre peut constituer un mode de balisage pour l'interprétation. L'hypothèse que je défends est que la situation applicative peut jouer le rôle de miroir grossissant de ce qu'est toujours une sémantique textuelle : la dimension interprétative (herméneutique pour reprendre un terme cher aux chercheurs en sémantique textuelle ((Adam, 1998), (Rastier, 2001)) y est toujours présente. Cette dimension, que l'on accepte assez facilement dans le cas d'une visée applicative, doit également être convoquée lorsqu'il s'agit d'une situation théorique.

En effet l'interprétation théorique s'inscrit elle aussi dans un point de vue sur les données, point de vue qui n'est pas neutre, et il est important que les paradigmes scientifiques qui nous influencent, dans lesquels nous nous inscrivons plus ou moins consciemment, soient clarifiés. Tout comme il est important que les demandes appliquées ne soient pas traitées au cas par cas mais considérées comme des instances d'une demande plus générale, que l'on peut exprimer sous la forme de questionnements théoriques. Ainsi, s'il existe des différences entre approche applicative et approche théorique, elles ne résident pas dans le fait que la première serait tributaire des résultats de la seconde. Les relations entre ces deux points de vue sont beaucoup plus complexes qu'il n'y paraît. Une des situations les plus fréquentes est celle dans laquelle la demande appliquée oblige à renouveler le questionnement sur cette problématique théorique³. Prenons un exemple rapide.

³ Précisons que cette demande « sociétale » n'émane pas toujours d'entreprises ; le besoin en traitement de données textuelles est très présent dès que la rédaction et l'exploitation de documents constitue une activité cruciale : cela peut être le cas d'organismes dépourvus de but lucratifs comme une association caritative ou un Organisme Non Gouvernemental. Que les entreprises soient plus particulièrement demandeuses, parce qu'elles ont les moyens de financer ces recherches et que les laboratoires de recherche aient besoin de ce mode de financement à la suite de choix politiques qui ne leur reviennent pas, n'a pas à voir, à mon sens avec la réalité de la problématique et la manière dont elle affecte la linguistique. La mutation est en lien avec la réalité du volume de documents à traiter et les possibilités ouvertes par le traitement automatique de la langue.

Une demande émanant du CNES a concerné la possibilité de mettre au jour des points de vue sur la base d'une étude de textes. Dans ce département de Recherche et Développement, la réflexion portait sur la conception (en l'occurrence de satellites) et sur le fait que chacun s'exprimait avec un point de vue, essentiellement celui de sa Division d'origine, mais sans avoir toujours conscience de l'influence que son appartenance pouvait avoir pour sa façon de concevoir un nouveau projet de satellite. Notre réflexion de linguistes (Condamines et Rebeyrolle, 1996) s'est déroulée en plusieurs temps. Elle nous a amenés à interpréter cette demande sur les points de vue comme relevant de la mise au jour d'un fonctionnement polysémique à partir de corpus. Le travail réalisé s'est fondé sur une analyse du mot *satellite* dans un corpus constitué de textes en provenance de deux divisions, Mathématiques spatiales et Observation de la terre, dans un premier temps non-distingués. Il s'est avéré que le mot *satellite* apparaissait dans 7 types de structures linguistiques qui ont pu être identifiés par des experts du CNES comme correspondants à autant de points de vue sur l'objet satellite (le satellite comme « objet à concevoir », comme « corps artificiel », comme « mobile », comme « plate-forme », comme « véhicule », comme « hôte », comme « relais »). Nous avons pu montrer qu'une structure apparaissait majoritairement dans chacun des deux sous-corpus (sans exclure complètement les autres structures), ce qui témoignait d'un point de vue dominant dans chacune des divisions d'où provenaient les corpus.

Du point de vue de la linguistique, repérer un fonctionnement polysémique sur la seule base d'analyse de données réelles est tout à fait nouveau. En général, ce qui est pratiqué, c'est la confirmation d'une polysémie supposée en langue par des analyses de textes. Dans ce cas précis, nous n'avions pas d'attente particulière sur *satellite* puisque nous n'avions pas la compétence linguistique sur le domaine spatial mais nous pensions bien que l'objet satellite lui-même était soumis à des points de vue différents. Notre réflexion s'est donc centrée sur les moyens de mettre au jour un fonctionnement polysémique, sur la base d'une analyse des contextes d'apparition d'un mot.

On le voit, une problématique initialement appliquée peut s'avérer riche en ouvertures théoriques. Mais alors, qu'est-ce qui différencie la situation appliquée de la situation théorique ? Deux éléments semblent pertinents pour essayer de caractériser cette différence :

- *Le problème de l'évaluation des résultats.* La sanction de la valeur des résultats obtenus par une analyse linguistique n'est pas la même selon que l'on se situe dans une perspective appliquée ou théorique. Dans une situation appliquée, la sanction revient à ceux qui ont émis un besoin particulier. Dans le cas de la situation

théorique, la sanction provient des pairs chercheurs qui évaluent la pertinence des résultats au regard de la connaissance de la discipline et des différents thèmes qui sous-tendent la réflexion à un moment donné. Dans les deux cas, la sanction elle-même est située, dans une histoire et dans un contexte donnés. Les deux types de sanctions ne sont pas incompatibles mais le mode de présentation des résultats doit être adapté aux évaluateurs, ce qui demande une réflexion non-négligeable, qui fait partie du travail de recherche.

- *Le problème de la réutilisation des résultats.* La réutilisation des résultats n'est pas forcément un enjeu majeur pour une situation appliquée. La demande est souvent ponctuelle et liée à un besoin immédiat. En revanche, pour des raisons financières, la tentation peut être grande pour un organisme de vouloir réutiliser des résultats dans un contexte qui sera considéré comme proche, souvent à tort. Ainsi, beaucoup d'entreprises souhaiteraient obtenir des ontologies (comme on dit en ingénierie des connaissances, cf. 3.1 et (Gomez-Perez, 1999)) qui seraient suffisamment générales pour pouvoir être utilisées dans un grand nombre d'applications : aide à la traduction, aide à la rédaction, systèmes de raisonnement, recherche d'informations... Mais, sur ce sujet en particulier, de nombreux travaux ont montré que ces applications étaient en réalité très variables et demandaient que soient construites à chaque fois de nouvelles ressources (Aussenac-Gilles, Condamines & Szulman, 2002). Pour une perspective théorique, la réutilisation des résultats est fondamentale car c'est ce critère de réutilisabilité qui garantit la systématicité. Pour l'analyse d'un corpus, cela signifie que l'on a pu dégager des régularités de fonctionnement suffisamment fortes, en lien avec les caractéristiques de ce corpus, pour que l'on puisse s'attendre à ce que ces mêmes régularités apparaissent dans un corpus ayant les mêmes caractéristiques. Par exemple, dans le cas des structures associées à *satellite*, on s'attend à ce que la structure linguistique caractéristique d'une Division se retrouve majoritairement dans un autre texte, relevant du même genre, émanant de cette même division. Cela suppose souvent de pratiquer une sorte de *méta-interprétation* qui a toujours pour objectif la mise au jour de régularités linguistiques en lien avec une régularité de la situation extra-linguistique.

3. Exemples de « demandes » et ouvertures théoriques

Lorsqu'un laboratoire de recherche est impliqué dans un projet avec une entreprise⁴, il est confronté à une double nécessité : d'une part prendre en compte un contexte particulier qui se traduit par une demande particulière, d'autre part, aller vers une théorisation des phénomènes, qui permet à la fois de s'inscrire dans une discipline et de réutiliser des résultats (en tout cas des méthodes) pour d'autres cas identiques ou similaires. Le chercheur sur le terrain se doit de se garder d'une vision rigide de sa discipline et de ses connaissances et doit faire preuve d'adaptation mais il doit aussi prendre garde à ne pas répondre de manière trop *ad hoc* à une demande précise : il n'est pas dans la vision d'un service à rendre moyennant financement. Cette nécessité est sans doute toujours présente lors d'intervention de chercheurs dans des entreprises mais elle prend une acuité particulière dans le cas de la linguistique dont la situation est fragilisée par le fait d'une arrivée très récente sur ce terrain.

Voici deux exemples de demandes qui nous ont été faites et des réponses que nous avons données.

3.1. L'exemple de la demande d'EDF

C'est le Département Qualité d'EDF (Clamart) qui nous a sollicités avec le problème suivant. Un manuel avait été rédigé sur la spécification en génie logiciel et la rédaction de documents en lien avec ces spécifications. Le service qualité souhaitait que ce manuel soit utilisé de manière systématique par les ingénieurs, ce qui était rarement le cas. La demande qui nous a été faite était de mettre au jour ce qui, dans ce manuel de 350 pages, pouvait être considéré comme l'information importante, qui pourrait être retenue et proposée sous un autre mode de présentation mieux accessible.

Nous avons opté pour une représentation relationnelle de l'information, représentation bien adaptée aux corpus techniques et sur laquelle nous avons déjà mené une réflexion. Cette représentation est couramment utilisée par différentes disciplines (terminologie (réseaux terminologiques), sciences de l'information (thesaurus), didactique (cartes conceptuelles

⁴ Dans le cas de l'ERSS, il s'agit dans la grande majorité de projets avec les départements Recherche et Développement de grands organismes publics ou semi-publics : CNES, EDF, EADS...

(Tiberghien, 1994), ingénierie des connaissances (ontologies)). Pourtant le passage d'une forme textuelle à une autre forme sémiotique a rarement été interrogée du point de vue sémantique. Brièvement le problème peut être représenté par la figure 1.

Il s'agit de passer d'une structure sémiotique discursive, linéaire et syntaxique à une structure volumique et iconique. Dans les deux cas, le code langagier est utilisé puisque les noeuds et les arcs de la structure relationnelle sont étiquetés par des éléments lexicaux, le plus souvent des noms pour les premiers et des verbes pour les seconds. L'interprétation consiste donc en une sélection de contextes dont on estime qu'ils peuvent être représentés sous forme relationnelle, puis en une catégorisation de ces passages pour repérer une relation similaire sous des formes différentes, enfin en un étiquetage sémantique de cette relation. Le repérage de ces passages pertinents se fait en utilisant ce qu'on appelle des *marqueurs de relations conceptuelles*, éléments lexico-syntaxiques permettant de repérer des relations sémantiques (ou conceptuelles). Certains de ces marqueurs sont bien connus et décrits dans la littérature ; par exemple, on sait que la structure `tous les N1 sauf le/la/les N2` (identifiée par introspection) permet de repérer une relation de générique à spécifique entre N1 et N2 (*Paul aime toutes les fleurs sauf les roses* permet de repérer une relation hiérarchique entre *rose* et *fleur*). Lorsqu'on se penche sur la réalité du fonctionnement des marqueurs en corpus, on se rend compte qu'il est beaucoup plus complexe que celui qui est généralement décrit. En effet, les marqueurs sont souvent présentés avec trois caractéristiques :

- il s'agit d'éléments très stables, non sensibles au contexte où ils apparaissent,
- il s'agit d'éléments que l'on peut identifier par introspection,
- il s'agit d'éléments qui donnent l'interprétation de la relation.

Or, l'analyse de différents corpus a montré que les marqueurs n'étaient pas fondamentalement différents d'autres signes linguistiques et qu'il était souvent inexact de leur attribuer un rôle d'indice de relation (Condamines, 2002).

En réalité, une façon de caractériser les marqueurs consiste à étudier leur lien de dépendance par rapport au corpus ; en partie grâce à l'étude menée sur le corpus d'EDF, trois types de fonctionnements ont ainsi pu être décrits :

- Le marqueur est indépendant du corpus. C'est le type de fonctionnement qui correspond aux marqueurs décrits par introspection. Un marqueur de relation générique-spécifique comme `Dét N1 être dét N2 + modifieur` (ex :

l'électron est une particule en mouvement) semble ainsi fonctionner de manière indépendante de la nature du corpus. Cela ne signifie pas que chaque fois que l'on va rencontrer ce marqueur, on trouvera un passage exprimable sous la forme relationnelle. Ce même marqueur peut être en effet utilisé pour exprimer une appréciation personnelle, par exemple, dans la phrase suivante extraite du roman *Germinal* :

Le brigand est le vrai héros, le vengeur populaire, le révolutionnaire en acte.

- Le marqueur est dépendant d'un genre de corpus. Certains marqueurs ne fonctionnent que dans un genre textuel bien déterminé. C'est le cas par exemple de *chez*, qui peut être associé à une relation de partie à tout dans un genre textuel bien précis : les textes didactiques dans le domaine des sciences naturelles (Condamines, 2000); ainsi dans :

Chez les colobinés, le nez fait saillie sur la lèvre supérieure,

il y a bien une relation de partie à tout entre *nez* et *colobinés*. Mais cette relation n'apparaît de façon massive (plus de 50% des attestations) que dans des textes du domaine des sciences naturelles, à condition qu'ils soient extraits d'un support didactique (encyclopédie, manuel...).

- Le marqueur est dépendant d'un corpus en particulier. C'est à l'occasion du travail sur le manuel d'EDF que nous avons rencontré un exemple évident de ce fonctionnement. En effet, dans ce corpus, la relation *conditionne le début de* se manifeste d'une façon tout à fait particulière (cf. Condamines et Rebeyrolle, 2000):

[[*(étape + phase) + déverbal*]] + (*lorsque, dès que*) + V au participe passé]

ex : *La phase d'intégration du composant peut commencer lorsque l'ensemble des éléments logiciels ont été codés.*

Dans ce corpus où le style injonctif est dissimulé sous une forme plus neutre, cette phrase doit en réalité être interprétée comme : *la phase d'intégration du composant ne peut commencer que lorsque l'ensemble des éléments logiciels ont été codés*, ou encore, *l'ensemble des éléments logiciels doivent avoir été codés avant le début de la phase d'intégration du composant*. Il s'agit donc bien d'exprimer une condition obligatoire. Ce mode d'expression, bien qu'interprétable par une analyse approfondie, était tout à fait impossible à identifier sur des bases introspectives (il aurait fallu penser à ce mode d'expression en réponse à la question « comment peut

s'exprimer la condition en discours ? »). Il s'agit donc d'un cas précis où l'analyse d'un corpus particulier a permis de mettre au jour un mode d'expression, certes propre à ce corpus mais qui relève d'une possibilité de la langue.

Les résultats que nous avons proposés à EDF ont permis au Service Qualité d'identifier quelles étaient les informations majeures qu'il souhaitait voir retenues par les ingénieurs. Ces informations ont été présentées dans un fascicule de quelques dizaines de pages dont la mise en forme a été confiée à des ergonomes.

3.2. L'exemple de la demande du CENA (Centre d'Etudes de la Navigation Aérienne)

Le projet avec le CENA a démarré à la suite d'une étude menée d'abord avec des ergonomes qui ont considéré avoir atteint les limites de leur compétence avec la demande formulée par le CENA⁵.

Le problème était le suivant, il s'agissait d'évaluer les éventuelles modifications introduites dans le mode de dialogue entre deux populations, à la suite de la mise en place d'une situation d'expérimentation simulée. Dans la situation réelle de dialogue, les protagonistes (un « chef de salle » et un « superviseur ») communiquent par téléphone à propos d'un dysfonctionnement technique d'un système, l'un des locuteurs exprimant les symptômes, l'autre devant interpréter ces symptômes. Mais, dans le cas de la situation simulée, les locuteurs peuvent visualiser des endroits de l'écran qui indiquent par exemple quel système est en panne. Cette situation simulée pouvait donc entraîner des modifications dans l'expression soit en raison de l'artificialité de la situation, soit en raison de la présence de cet écran partagé.

D'un point de vue plus théorique, l'étude⁶, encore en cours, se concentre sur l'expression du dysfonctionnement technique et vise à évaluer si des modification extra-linguistiques, quelles qu'elles soient, ont une influence sur ce mode d'expression et, si oui, quelle est la nature de ces modifications. Trois corpus ont été constitués en fonction de deux types de variations du contexte extra-linguistique : la période de production et la nature simulée ou réelle de la situation :

- corpus réel, 1995,

⁵ Il serait d'ailleurs tout à fait nécessaire que les complémentarités entre ergonomie et linguistique sur l'analyse de textes soient étudiées en profondeur. Il n'est pas rare en effet que l'on sollicite des ergonomes, souvent déjà présents dans les entreprises, pour des études pour lesquelles les linguistes seraient mieux armés. L'inverse n'est pour l'instant pas vrai faute de linguistes en nombre important sur le terrain mais il est évident que la situation serait plus profitable à chacun si les contours des compétences étaient au moins interrogés.

- corpus réel, 2001,
- corpus simulé, 2001.

Le premier corpus a servi de corpus de référence. L'objectif de l'étude a consisté à essayer de repérer dans ce premier corpus des régularités dans les énoncés du dysfonctionnement, c'est-à-dire de définir une sorte de grammaire de l'expression du dysfonctionnement qui prenne en compte tous les énoncés et, à partir d'eux, élabore des structures syntaxiques modélisant ces énoncés. Ces structures permettent de localiser l'élément dysfonctionnant et la nature du dysfonctionnement. L'objectif dans un second temps est de comparer les corpus en appuyant l'étude sur la grammaire ainsi élaborée.

Les structures construites sont au nombre de trois :

- Structures « assertives » : Sujet + prédicat :

Ex : *La fréquence 118.85 est brouillée par une fréquence heu, anglaise.*

L'élément dysfonctionnant est *la fréquence 118,85* et la nature du dysfonctionnement est *être brouillée par une fréquence anglaise*.

- Structures « présentatives » : Présentatif + Syntagme Nominal + (proposition relative, Syntagme prépositionnel, Syntagme Nominal) :

Ex : *C'est le STPV qui est planté* (élément dysfonctionnant : *STPV*, dysfonctionnement : *être planté*).

J'ai un digit qui est en panne (élément dysfonctionnant : *digit*, dysfonctionnement : *être en panne*).

- Constructions « détachées » : (Syntagme prépositionnel, Syntagme Nominal) + phrase :

Ex : *A propos de la 130.95 elle n'est plus utilisable* (élément dysfonctionnant : *130.95*, dysfonctionnement : *ne plus être utilisable*).

Les premiers résultats de comparaison montrent une grande stabilité des résultats entre le corpus d'étude et les deux autres. Il existe des différences mais elles ne concernent pas les structures mises en œuvre elles-mêmes mais plutôt le vocabulaire utilisé. En d'autres termes, les modes d'expression du dysfonctionnement technique ne semblent pas être affectés par des changements dans la situation de production de ces énoncés (rappelons qu'il s'agit d'une

⁶ Cette étude est menée par P.Vergely dans le cadre de son doctorat en sciences du langage (Vergely, 2002).

situation de dialogue, les structures mise en oeuvre dans un texte écrit sont certainement d'une autre nature). Cette stabilité des structures permet d'effectuer une comparaison des corpus qui met en évidence rapidement les changements d'un corpus à l'autre. Outre la possibilité d'asseoir les observations sur des éléments stables, cette étude a un autre intérêt, majeur dans la perspective d'une linguistique de corpus. En effet, il se peut que les régularités dans l'expression du dysfonctionnement ne soient pas propres au CENA ni même au domaine de l'aéronautique ni même à quelque domaine que ce soit. On peut faire l'hypothèse que la langue, en tant que système, met à disposition certaines structures, que nous utilisons inconsciemment de manière régulière lorsque, dans le cadre d'un dialogue, nous devons exprimer un dysfonctionnement. Cette hypothèse devrait être confirmée par une exploration de corpus en provenance d'autres domaines, par exemple le contrôle maritime et plus seulement le contrôle aérien, mais aussi le domaine médical : expression de symptômes physiques, dans le cadre d'une visite médicale par exemple.

Si ces régularités se révélaient aussi stables que nous le pensons, elles pourraient servir d'argument à la possibilité d'une analyse linguistique ascendante, des corpus vers le système linguistique, ce qui serait une façon de concilier approche sur corpus et description du système de la langue. Jusqu'à présent en effet, les corpus ont surtout été étudiés pour eux-mêmes, pour mettre en évidence leur spécificité. Examinés dans la perspective de décrire la langue, ils prendraient un tout autre relief. Et la description de ce qui fait système serait elle aussi à réévaluer. Il ne s'agirait plus d'étudier le fonctionnement de telle ou telle catégorie grammaticale (les prépositions, les verbes...) mais d'orienter l'étude sur des noyaux de sens constitués et sur le rôle des éléments structurels pour les exprimer : par exemple ici, le dysfonctionnement, et, pour l'exemple du paragraphe 3.1, l'expression de telle ou telle relation. La variation passerait alors majoritairement par le genre du corpus (sciences naturelles et didactique pour le cas de l'expression de la partie-de avec *chez*, dialogue entre un expert et un moins expert pour les structures syntaxiques de l'expression du dysfonctionnement).

4. Conclusion

Du fait de la mise à disposition de très nombreux textes via l'internet ou via les intranets, la linguistique ne peut plus se tenir à l'écart d'une prise en compte de corpus réels. La production de ces corpus est toujours située : il ne s'agit pas de monceaux d'attestations dans lesquels on peut puiser ce qui vient confirmer des intuitions. Une des manifestations de la

réalité de ces usages est donc liée à la réalité des situations dans lesquelles ils sont produits. L'autre réalité, qui peut être difficilement acceptable pour un linguiste imprégné de l'idée que sa discipline relève d'une science quasi exacte, vient du constat incontournable que l'analyse d'un texte ne relève pas de la simple découverte du sens de ce texte mais bien de la construction d'un sens et donc d'une interprétation des faits langagiers. L'objectif de l'analyse joue donc un rôle majeur dans la production des résultats. Une façon d'essayer de comprendre comment s'élabore l'interprétation pour évaluer ce qui peut être systématisé consiste à contrôler cette interprétation en caractérisant très précisément deux situations : la situation de production des textes du corpus et la situation d'analyse (objectif). Dans cette perspective, la situation appliquée, loin de n'être que la mise en œuvre de résultats de l'analyse théorique, peut contribuer à alimenter l'analyse théorique voire à lui donner de nouvelles perspectives. Il n'y a donc pas de sens à opposer analyse théorique et analyse appliquée ; il faudrait au contraire essayer de voir comment les deux situations d'interprétations peuvent se compléter et s'éclairer. Pour ôter de la dimension applicative l'idée péjorative qu'elle contient du fait de son histoire, certains auteurs préfèrent parler de linguistique impliquée (Léglise, 2000), cet épithète est effectivement bienvenu parce qu'il prend en compte et valorise le rôle social du linguiste qui avait été longtemps minimisé à l'avantage d'un point de vue essentiellement théorique du fonctionnement de la langue. Ce changement correspond certainement à une évolution épistémologique majeure pour la linguistique

Bibliographie

- ADAM, J.-M. (1999) *Linguistique textuelle. Des genres de discours aux textes*. Paris : Nathan.
- AUROUX, S. (1998) : La raison, le langage et les normes. Paris : PUF.
- AUSSENAC-GILLES, N., CONDAMINES A., & SZULMAN S. (2002) Prise en compte de l'application dans la constitution de produits terminologiques . *Actes des 2e Assises Nationales du GDR I3*. Nancy (F), Déc. 2002. pp 289-302. Toulouse : Cépaduès Editions.
- BOUTET, J. (1995) *Construire le sens*. Bern, Berlin : Peter Lang. 1994.
- CONDAMINES, A. (2000) Chez dans un corpus de sciences naturelles : un marqueur de méronymie ? . pp. 165-187. *Cahiers de Lexicologie* n° 77.
- CONDAMINES, A. (2002) Corpus Analysis and Conceptual Relation Patterns . pp. 141-162. *Terminology*, volume 8 number 1.
- CONDAMINES, A., & REBEYROLLE, J. (1996) Point de vue en langue spécialisée. pp. 174-184. *META*. 42, 1.
- CONDAMINES, A., & REBEYROLLE, J. (2000) Construction d'une base de connaissances terminologiques à partir de textes : expérimentation et définition d'une méthod. pp. 225-242. J. Charlet, M. Zacklad, G. Kassel & D. Bourigault, (eds). : *Ingénierie des Connaissances, évolutions récentes et nouveaux défis*. Paris : Eyrolles.

- CORBIN, P. (1980) De la production des données en linguistique introspective. pp. 121-179. A.M. Desseaux Berthoneau (ed) : *Théories linguistiques et traditions grammaticales*. Lille : PUL.
- DUBOIS J., GUESPIN, L., GIACOMO, M., MARCELLESI, C., MARCELLESI, J.B., & MEVEL, J.P. (1973) *Dictionnaire de linguistique*. Paris : Larousse.
- FOUCAULT, M. (1966) *Les mots et les choses*. Paris : Tel, Gallimard.
- GOMEZ-PERREZ, A (1999) Développements récents en matière de conception, de maintenance et d'utilisation des ontologies ». pp. 9-20. *Terminologies Nouvelles* n°19.
- HABERT B., NAZARENKO, A., & SALEM, A. (1997 : *Les linguistiques de corpus*. Paris : Armand Colin.
- LABOV, W. (1976) *Sociolinguistique*. Paris : Editions de Minuit.
- LEGLISE, I. (2000) Lorsque les linguistes interviennent : écueils et enjeux. pp. 5-14. *Revue Française de Linguistique Appliquée*. Volume V-1, juin 2000.
- PERY-WOODLEY, M.-P. (1995) Quels corpus pour quels traitements automatiques. pp. 213-232. *TAL* (Traitement Automatique des Langues), n°36 (1-2).
- PIERREL, J.-M. (2000) *Ingénierie des langues*, Traité I2C, Paris : Hermes.
- RASTIER, F. (2001) *Arts et Sciences du texte*. Paris : PUF, formes sémiotiques.
- TIBERGHEN, A. (1994) Choix sous-jacents à la construction de représentation spatiale de concepts. pp. 53-62. *Didaskalia* n°5. Décembre 1994.
- VERGELY, P. (2002) Régularités linguistiques d'un langage opératif : le cas de la NavigationAérienne. pp. 59-70. *Actes du GLAT (Groupe de Linguistique Appliquée des Télécommunications)*. Mai 2002.

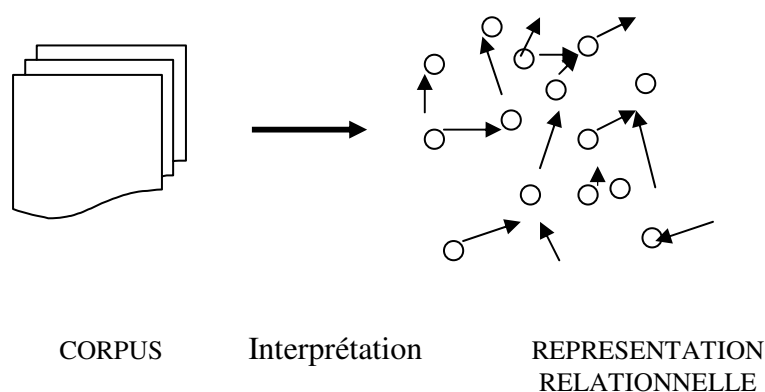


Figure 1 : d'un corpus à un réseau relationnel